The A.I.D. Evaluation System: Past Performance and Future Directions

September 1990

Bureau for Program And Policy Coordination
Agency for International Development

TABLE OF CONTENTS

FOREWORD

In a recent memorandum, the A.I.D. Administrator, Ronald W. Roskens, tasked the Bureau for Program and Policy Review with the preparation of a "think piece" on the subject of evaluation. He expressed his concern that the Agency's evaluation system must be able to assess the impacts of its programs with some degree of assurance and provide operationally useful lessons from experience to guide future program planning and policy decisions. The Administrator asked that options for strengthening the evaluation system be examined in order to make it more useful at all Agency management levels, from senior management to the project operating officers.

Sharing the Administrator's concerns about the importance of this topic, I instructed the Center for Development Information and Evaluation (CDIE) to undertake a serious and thorough assessment of how evaluation has been practiced in A.I.D. Working closely with CDIE, I requested that the review address the following major questions:

1. What are the major legitimate purposes and audiences for evaluation in an agency such as A.I.D.? What are their information needs?

2. What are the implications of these diverse purposes, audiences and information needs for selection of appropriate evaluation methods and procedures? For the optimal organization of evaluation responsibilities?

3. How has A.I.D.'s actual evaluation system evolved, and how is it organized, staffed, and funded today?

4. Against this background, what are the problems and strengths of the evaluation system as practiced today in A.I.D.?

5. Based on this assessment, what are some of the options and suggestions for strengthening the A.I.D. evaluation system?

I have been particularly concerned about improving the ability of the current evaluation system to meet senior management's information needs, and to implement more rigorous, comparative and objective evaluations than we have been able to do in the past. This paper, drafted by Annette Binnendijk, Chief of the Program and Policy Evaluation Division, develops a framework for considering these and other issues, and should serve as a basis for Agency-wide discussions on strengthening evaluation in the Agency.

Reginald J. Brown
Assistant Administrator
Bureau for Program and Policy Coordination

INTRODUCTION

In light of the continuing debate about A.I.D.'s effectiveness, and Congressional interest in reporting on development results, the Bureau for Program and Policy Coordination has decided to undertake a thorough review of the Agency's evaluation system. This paper was prepared to lay the groundwork for such a review. Its purpose is to assess the performance of the existing evaluation system, set it in the context of what it should be doing, and make suggestions for improvements.

The paper is organized into five sections. The first two sections develop a conceptual framework identifying the potential purposes, audiences, and information needs that A.I.D. evaluations might ideally address, and the implications for selection of appropriate evaluation methods, practices, and organizational structures. The next section turns to look at actual evaluation practice, first examining how the evaluation system evolved in A.I.D. and the forces that helped to shape it, and then reviewing how the system operates today, in terms of allocation of evaluation functions, responsibilities, staff and resources among various organizational levels. The fourth section assesses the A.I.D. evaluation system's progress at the different organizational levels, and identifies significant remaining problems and their probable causes. The critiques are made in light of the theoretical framework developed earlier, by comparing "what is" with "what could be". The final section concludes with a number of suggestions for strengthening evaluation in A.I.D.

I. THE PURPOSES AND AUDIENCES FOR A.I.D. EVALUATIONS

Before assessing the effectiveness of A.I.D.'s existing evaluation system, it is useful to think more conceptually about what the purposes of evaluation might be in an agency such as A.I.D., and about who are legitimate audiences and what are their special information needs.

In the private sector, firms get clear and simple indications of their performance from market prices, and the "bottom line" of profits and loss. However, in public bureaucracies such as A.I.D., assessment of an organization's performance is more complex. "Market tests" are replaced by more elaborate evaluation methodologies. Government agencies are accountable to the Congress and constituencies for achieving results for funds spent, and thus evaluations must measure program performance and efficiency. Also, evaluations should provide useful information to assist the agency's managers in making programming and policy decisions that improve achievement of the organization's objectives.

Management theory stresses that the most effective organizations are those in which management keeps clear sights on its long-range or ultimate goals and sets organizational priorities, values, and actions to achieve this vision. In the case of A.I.D., the ultimate mission is achieving meaningful development results, such as promoting broad-based, market-oriented and sustainable growth while enhancing human dignity, freedom, and quality of life around the world. In this context, evaluation can be viewed as a critical "tool" that management has to assess whether development objectives are being achieved, and how best to achieve them. However, in A.I.D.'s decentralized organizational structure, the concept of evaluation serving as a management tool is complicated by the multiple levels of management and the varying nature of their information needs.

Evaluation could serve several major, legitimate purposes and audiences in A.I.D.:

1. Evaluations should support operational decisions at the level of individual interventions, for example, for projects, non-project assistance, and other modes of assistance. The primary audiences and users of this type of evaluation include the A.I.D. mission project managers and their host country counterparts in implementing agencies. Their specific information needs include:

(a) monitoring of implementation processes, such as the delivery of various project inputs and outputs according to design targets and uncovering various sources of implementation delays and shortfalls.

(b) assessments of project performance or effectiveness in achieving intended project purposes and intermediate results, for example, whether a project's services are being used appropriately by the intended target groups. Such information can serve project management's decision needs for redirecting implementation efforts, correcting problems, and undertaking mid-course redesign efforts that may be necessary to help ensure that intended purposes and results will be achieved.

2. Evaluations should inform and influence major Agency policy and programming decisions. Major audiences would include senior management in A.I.D./W (Administrator's office, central and regional bureau AAs, DAAs, and office directors), USAID mission directors, and host country leaders. Examples of their specific information needs include:

(a) comparative assessments that help formulate Agency policy and program guidance statements, and project design guidelines, by drawing lessons from experience about what intervention and policy approaches work best in what country conditions.

(b) comparative assessments for resource allocation decisions--for example at the mission level, information on program performance relevant to decisions concerning which interventions to support, which to revise, and which to terminate. At more senior A.I.D./W levels, the information would need to be relevant to resource allocation decisions among countries and regions, and among competing, alternative intervention approaches, based on their performance in achieving results cost-effectively.

3. Evaluations should support Agency management's "accountability" responsibility by explaining the nature, performance, and impacts of the Agency's assistance efforts to external oversight and constituency groups. Key audiences would include the U.S. Congress, OMB, GAO, the State Department, various special interest and constituency groups, the press, and the general American public. The accountability concept might also be extended to include A.I.D.'s recipient countries. Information needs include:

(a) reports on the nature, character, and magnitude of the assistance program

(b) quantitative information on the performance, efficiency, and developmental impacts of major Agency programs aggregated at country and regional levels

(c) increasingly, evidence that the Agency's management is being guided in its operational decisions by a concern for achieving development results cost-effectively; that is, following a results-oriented, performance-based system of management.

4. Some might argue that A.I.D. evaluations should also contribute to the broader knowledge base about the development process and the role of donor agency interventions. Audiences might include other bilateral and multilateral donors, international and regional development organizations, developing country governments, PVOs, NGOs, academic institutions, and others in the development community. The information needs of such diverse audiences are probably very broad (and would be very difficult to try to address directly), but might include information and evidence of what types of donor agency intervention and policy approaches work under various development conditions, as guidance for future development strategies. Here, we assume that while legitimate, this purpose is less critical or secondary to the others, and can be treated as a "by-product" of the other purposes. That is, evaluation studies should be broadly disseminated or accessible to such audiences, but in general should not be specially undertaken to meet their specific needs.

II. SELECTING APPROPRIATE EVALUATION METHODOLOGIES AND PROCEDURES

This diversity in potential, legitimate purposes, audiences and information needs for A.I.D. evaluations, has implications for the types of evaluation methodologies and practices that would be most suitable. Unfortunately, it cannot be assumed that any single evaluation approach will meet the needs of all Agency management levels and all audiences. In fact, there appears to be some very real differences in terms of most appropriate methods, procedures and structuring of organizational responsibilities for evaluation, depending on major purposes. In particular, the needs of the project officer in the field dictates a type of evaluation methodology and approach which may be quite different from the needs of A.I.D. senior management or of the Congress.

A. ISSUES IN THE SELECTION OF EVALUATION METHODS

Before proceeding to discuss the most appropriate methods for different categories of A.I.D. evaluation purpose and audience, it will be useful to discuss briefly several aspects or dimensions of evaluation theory and practice, as it relates to the A.I.D. experience.

Quantitative versus Qualitative Methods

Some A.I.D. evaluations use quantitative or statistical methods, while others use more qualitative, or rapid appraisal techniques. Some combine approaches usefully. Systematic use of innovative qualitative techniques has proven to be very useful for some evaluation purposes, especially where low-cost approaches and quick results are needed. These methods include focus group interviews, village or community meetings, key informant interviews, site observation techniques, and small informal surveys. However, they will not serve some evaluation needs, especially where more rigorous evidence of impacts is needed or where there is a need to compare or aggregate performance or impact data across countries, regions, projects or programs.

Technical versus Practical Criteria Tradeoffs

The decision about what type of evaluation methods and indicators to use can also be viewed in the context of tradeoffs between a variety of technical and practical criteria. For example, in an ideal world where practical considerations of time, cost, technical capabilities of staff, etc. were not constraints, one would ideally select evaluation indicators and methods that had the greatest validity, reliability, and comparability. In other words, the most rigorous, objective, and convincing methods available. However, we know that tradeoffs exist; for example, the more rigorous the approach, the more time consuming, expensive, and complex the effort is likely to be.

Internal versus External Evaluations

Another issue in the selection of evaluation methods is the tradeoffs between choosing an "internal" versus an "external" evaluation approach. In an external approach, emphasis is on ensuring objectivity and reducing possible biases by forming an evaluation team composed of individuals far removed from the project operations and thus less likely to have vested interests in the evaluation findings. In an internal approach, emphasis is instead on ensuring greater utilization of evaluation findings by including project management and other stakeholders in the process. The relative importance of these criteria, and thus the appropriateness of the approach selected, will depend in part on the purposes and audiences for which the evaluation is undertaken. For example, if the primary purpose of an evaluation is to guide operational decisions of project management, an internal or stakeholder approach is probably preferable. On the other hand, if the evaluation is to be used for purposes such as resource allocation decisions or for external accountability, the more objective external approach is indicated.

Evaluation Foci

Evaluations of A.I.D. interventions may focus on many levels or issues, which might be conceptualized as falling into the following categories: (a) implementation monitoring; (b) effectiveness evaluation; (c) impact evaluation; (d) efficiency evaluation; and (e) sustainability assessments. The major differences between these types of evaluation can be most easily explained by thinking of the levels of the project Logframe.

- monitoring is concerned with assessing "inputs" and "outputs"
- effectiveness evaluation is concerned with performance in achieving project "purposes" or intermediate results
- impact evaluation focuses on achievement of project "goals" or ultimate long-term results
- efficiency evaluation examines project costs in relation to performance and impacts (e.g., cost-effectiveness, cost-benefit analysis)
- sustainability evaluation looks at the issue of continuing project results and impacts after the project is completed and donor funding ends.

The issues on which a particular evaluation focuses has implications for the selection of appropriate indicators and methodologies, as discussed in appendix A.

Project- versus Program-Level Evaluations

Another issue in selecting appropriate evaluation methods is whether interest is at the individual intervention level or "above," at a more programmatic or sectoral level. Most conventional evaluation methods have focused on assessing results of specific interventions, primarily the project approach. However, in recent years, A.I.D.'s development strategy has shifted away from exclusive reliance on projects to include other forms of non-project assistance. Also, Congressional and A.I.D. senior management interests are increasingly focusing on assessing what the overall strategic results of our assistance have been on broader macro-level or program goals--that is, on whether the Agency is making a significant difference in developing economies and societies.

Such concerns call for more comparative types of evaluations, capable of aggregating, synthesizing, and comparing results across individual interventions. Increasingly, A.I.D. is interested in "program performance" evaluations, that is, assessments of how various interventions (projects, policy reforms, food aid, etc.) may have contributed to a particular strategic or "program" goal (e.g., promoting private sector development, hunger alleviation, child survival). Pilot efforts are now underway in several missions to develop appropriate methodologies and indicators for such program performance evaluation systems focused on assessing A.I.D.'s progress towards achieving strategic objectives in particular countries.

In addition, A.I.D.'s central evaluation office, CDIE, and the regional bureaus, have developed and used mostly qualitative "rapid appraisal" methodologies for synthesizing Agency-wide or region-wide findings and lessons at the program level, based on document reviews or field evaluations of a series of interventions within a program area.

B. RELATIONSHIPS BETWEEN EVALUATION PURPOSES AND APPROPRIATE METHODS AND PRACTICES

As discussed earlier, the multiple potential purposes and audiences for A.I.D. evaluations, and their varying information needs, have implications for the selection of appropriate methods, approaches, and processes, and for the structuring of organizational responsibilities of the evaluation function. **The various levels of potentially legitimate uses and users of evaluation information in A.I.D. suggest a need for a variety of methods and distribution of organizational responsibilities suited to meet these diverse needs.** Getting to specifics, the discussion that follows examines the characteristics of evaluation methods and approaches that seem most suited for each major purpose. To simplify the presentation, some of the dichotomies presented below may be somewhat overstated. (See Figure 1 for a summary presentation).

1. Support for project management's operational decision needs implies evaluation methods and approaches that:

- provide rapid appraisal and feedback to inform implementation and redesign efforts for ongoing projects
- are relatively low cost since they should be undertaken regularly and for all mission interventions
- are highly collaborative in nature, involving host country counterparts in the process
- require relatively simple technical capabilities and skills to undertake
- are project-specific and directly relevant to project management's operational decision needs
- emphasize monitoring of implementation progress and problems, and tracking of shorter term results via "intermediate" indicators of project effectiveness and performance, such as access to and utilization of services.

Evaluation methods for this purpose level need not necessarily be:

- overly concerned with measuring longer term, ultimate impacts
- highly rigorous or quantitative, or comparable with other projects in other countries or sectors
- designed to ensure a high level of objectivity or independence; since the sponsor of the evaluation is the primary user, "bias" is not a great issue for such evaluations.

Organizationally, if past experience is any guide, this type of project-level monitoring and evaluation function is best sponsored and managed directly within the project management structure. To be directly relevant and utilized by project management for operational decisions, the organizational unit responsible for such ongoing evaluation is best positioned as a project component and should be planned, funded, and staffed under the project design.

2. Information to guide major Agency programming and policy decisions implies evaluation methods and approaches that:

- examine the longer term, more ultimate development results and impacts of our interventions. Thus, the methods selected need to deal effectively with the issue of attributing changes to specific interventions.
- are relatively more rigorous, quantitative, and comparative in nature. Especially information needs for resource allocation decisions imply the use of more statistical and standardized indicators and methods for comparative evaluations across projects, programs and countries, e.g. cost-effectiveness analysis, cost-benefit or economic rates of return, when appropriate and feasible.
- ensure objectivity of evaluations and guard against bias and vested interests, especially if uses of the information involve resource allocation decisions. This implies that "external" evaluators need to be involved with independent authority to avoid the potential for bias that might occur if evaluation

Purposes	Audiences	Information Needs	Appropriate Methods and Procedures
1. Support Operational decisions at the level of individual development interventions	• A.I.D. mission project managers • host country project counterparts	• monitoring of implementation processes • preliminary assessments of project performance or effectiveness in achieving intended purposes and intermediate results	• rapid, low cost • qualitative • collaborative • require simple technical skills • intervention level • comprehensive of all interventions • focused on operational issues • managed "internally" by project management
2. Inform and influence major Agency policy and programming decisions	• AID/W senior management • USAID mission directors • host country leaders	• comparative assessments for resource allocations decisions • comparative assessments for A.I.D. policy and program guidance and for project design guidelines	• long term and rapid response • quantitative & qualitative • more rigorous, comparative • require sophisticated technical skills • focus on program performance and impacts • program level • representative sample of intervention • objective, independent • centrally (AID/W) or mission sponsored
3. Support Agency management's accountability responsibility to external constituencies	• U.S. Congress • OMB • GAO • State Department • other special interest and constituency groups • the press • American public • recipient governments	• reports on the nature, character, and magnitude of the aid program • convincing evidence about achievement of development results of A.I.D's programs • evidence of a performance-based management system in A.I.D.	• long term, costly • focus on program performance & impacts • quantitative, rigorous, comparative • program level • comprehensive • understandable to external audiences • objective, independent • centrally (AID/W) sponsored studies, or central reviews/audits of mission reporting
4. Contribute to broader knowledge base about the development process	• other donors • international and regional development organizations • developing countries • PVO's and NGO's • academic institutions • others in the development community	• very broad — difficult to address directly	• broadly distribute evaluation studies prepared to meet other purposes

is left up to those who would gain from good performance results.

- serve urgent information needs with rapid appraisal techniques while using more rigorous, statistical methods where there's sufficient time for longer term study and enough sustained interest to justify the expense.

Evaluation methods for this purpose need not necessarily be:

- limited to simple techniques or lower level technical skills
- comprehensive, i.e., cover the whole program portfolio. A carefully selected sample of interventions may be used to draw broader conclusions
- always constrained by needs for quick results or low-cost approaches. While senior management will certainly have considerable needs for low-cost and rapid response appraisals for specific and urgent decision needs, other studies in continuing high priority areas could afford more time and money for more rigorous evaluation efforts.

Organizationally, this type of evaluation purpose and approach requires central (or regional) sponsorship and control of the evaluation process, given the need for standardization of indicators and methods, the need for objectivity, etc. It implies a well-trained "core" staff of evaluators who have authority and independence to be objective in their findings, and sufficient resources and time to implement more rigorous, sophisticated evaluation methodologies. Perhaps mission evaluation reporting systems could be used to a limited extent, but there would have to be some central guidance and review/audit effort to attempt to ensure objectivity and comparability of mission program performance and evaluation reporting systems.

3. Support for the external accountability function requires evaluation methods that:

- are convincing, credible, and understandable to the external audiences
- are focused on higher level program performance and impacts
- go beyond isolated "success stories" to provide more quantitative evidence of overall program progress in achieving key Agency goals. This implies the need for more objective, rigorous and comparative methods that can aggregate or sum up performance and impact indicators from individual interventions up to country, regional, and perhaps world-wide levels.
- ensure independence and objectivity of findings
- are relatively comprehensive of the Agency portfolio, if the task is to assess the sum total results of funds spent (in a particular program, country, region, etc.).

In general, evaluations for this purpose need not be:

- quick turn around or low cost
- focused on project-level operational decision needs

Organizational implications for undertaking this type of evaluation function would include the need for a central (or regional) professional staff responsible for doing "accountability" analysis and for reporting to external audiences. Because of considerable similarities between the evaluation methods and approaches required to satisfy internal senior management needs and external accountability audiences, the same organizational units could be tasked with both functions. A centralized reporting system based upon missions' evaluation systems might be used for this purpose, provided central guidance could ensure standard indicators for comparing and aggregating results, and sufficient oversight to ensure objectivity.

III. EVALUATION PRACTICE IN A.I.D.

Previous sections have laid out a theoretical framework of what evaluation methods, practices, and organizational structures might best serve several legitimate purposes and audiences of A.I.D. evaluations. Now, we turn to look at actual evaluation practice--first examining how the evaluation system evolved historically in A.I.D. and the forces that helped to shape it, and then reviewing in some detail how the system operates today, in terms of evaluation functions and responsibilities, staff, and resources allocated to various organizational levels.

A. A BRIEF HISTORY OF THE EVOLUTION OF EVALUATION IN A.I.D.

Major changes in A.I.D.'s evaluation system over the years reflect broader Agency programming and policy shifts, increasing delegations of authority to the field missions, and responses to experiences with various evaluation approaches. For example, evaluation methodologies had to adjust to shifts from capital and infrastructure projects of the 1960s, to the basic human needs projects of the 1970s, and then to the greater emphasis on non-project assistance and macro-economic policy reform efforts of the 1980s. Evaluation system orientation, procedures, and organizational arrangements also shifted beginning in the early 1980s in response to the overall Agency trends towards decentralization. Furthermore, practical and methodological difficulties encountered in attempting rigorous, statistical approaches to evaluation in the 1970s led to greater emphasis on more qualitative approaches in the 1980s. Some of these major changes in orientation are discussed below.

During the 1960s, when most development projects were large capital (infrastructure and industrial) projects aimed at maximizing economic growth, the most popular appraisal and evaluation approach was estimating financial and economic rates of return. However, by the early 1970s, A.I.D.'s program emphasis shifted to technical assistance and institution building approaches, and to meeting the basic human needs of the poor through social services and agricultural and rural development projects focused on assisting impoverished groups. As a consequence, the evaluation technique of estimating financial and economic rates of return became increasingly insufficient.

Problems with the methodology included its inability to deal adequately with the equity and distribution issue, and the difficulties involved in trying to assign monetary value to all of a project's social benefits and other impacts.

What emerged in A.I.D. in the 1970s was a conceptual framework for guiding project design, implementation, and evaluation efforts known as the "Logical Framework" that well-suited the basic human needs project approach to development. The A.I.D. evaluation guidance of the early 1970s that accompanied the Logframe emphasized evaluation designs based on quasi-experimental designs and statistical controls, whereby impacts, or changes in the living standards and behaviors of the project beneficiaries, could be measured and held attributable to project interventions. This methodology required that statistically representative sample surveys be taken as a baseline followed periodically by other surveys over the project's lifetime in order to track changes in outcomes. The professional evaluation community and other donors were also advocating this approach to the evaluation of development projects during this time.

However, this evaluation approach proved to be very difficult to implement, requiring specialized skills, years to complete, and considerable expense. Ultimately it suffered from methodological weaknesses, inconclusiveness, and an orientation of little practical usefulness or interest to project managers. By the early 1980s, the popularity of this approach was declining in A.I.D. (and more generally among professional evaluators and other donors). In fact, it had never really been widely used by A.I.D. at any rate. There was a methodological gap. For some years, there were no practical, low-cost evaluation techniques offered as alternatives. Little empirical information was being gathered systematically on beneficiaries' use of project outputs, technologies, and services or their impacts. Most mission "evaluations" were really monitoring reports.

Other trends in evaluation approaches reflected the Agency's general shift to a more decentralized management system, with greater delegation of authority to the field. In the 1970s, A.I.D.'s evaluation system was relatively centralized, as was the Agency's management structure. A.I.D./W required that all projects be evaluated annually according to a rather rigid reporting format and set of questions that followed the Logframe concepts. However, evaluation reporting was often seen as an empty "paper" exercise, with too little empirical data being collected. Moreover, officers in the field tended to see the evaluation exercise as an "imposed" A.I.D./W requirement with little or no operational relevance to them or their projects. Little use was made by A.I.D./W of the mission evaluations sent from the field; few attempts were made to synthesize findings or extract lessons relevant for future Agency program or policy decisions.

Another problem was that projects typically lacked monitoring and evaluation units to provide ongoing collection of evaluation data as part of the project process. Emphasis was on "external" evaluation teams flying in and out, which proved to be inadequate for meeting information needs for operational decisions. There was a need for more permanent, on-site, and well-planned efforts with adequate resources within projects which would be responsive to project management's information needs.

By the early 1980s, several shifts occurred in the A.I.D. evaluation system and procedures, reflecting greater Agency decentralization and delegation of authority to the missions in general, as well as incorporating lessons from previous experience with evaluation. The changes also reflected new concepts and methods being advocated in the professional evaluation community, such as utilization-oriented evaluations and stakeholder involvement. In 1980, the universal, annual requirements for evaluation reporting to A.I.D./W were dropped. By the mid-1980s, central and regional bureau guidance to the missions were advocating the following new evaluation philosophy, procedures, and methods:

- A new emphasis was placed on making evaluations relevant and useful to project management's operational needs. New guidance stressed the principle of management's utilization of evaluation findings.

- The guidance also emphasized that management should play an active role in the evaluation process if the evaluation system was to be successfully utilized. Asking the evaluation questions to be examined and reviewing and following up on evaluation recommendations were viewed as critical management tasks.

- A new evaluation reporting format, more flexible than the old, allowed management in the field to ask the relevant evaluation questions for their own practical decision-making needs.

- The new procedures advocated that special evaluation studies be undertaken selectively at the initiative of various management levels, and timed to meet their needs for information to make critical operational decisions.

- An Agency annual evaluation planning process was established that was to coordinate evaluation planning and needs at the Agency's various management levels. The process was to start at the project level, working up to mission, then regional bureau, and finally Agency-wide level. At each stage, there would be coordination to ensure that the evaluation information needs at all management levels were being incorporated, priorities established, duplications avoided, etc.

- Guidance recommended that project designs build in ongoing monitoring and evaluation systems as integral project components, with adequate staff and resources for gathering results-oriented performance data on a regular basis.

- Collaborative evaluation efforts involving host country counterparts were strongly advocated.

- Promising innovative methodologies were being developed, explained, and encouraged in guidance reports that were beginning to fill the "methodological gap." These techniques included key informant interviews, focus group discussions, community meetings, observation methods, and small,informal surveys. These mostly qualitative methods were more practical, low cost, timely, and relevant for project management's information needs than were the more statistically rigorous methods advocated earlier. Yet they enabled systematic and empirical investigation into development results and impacts. Statistical rigor in evaluation methods was deemphasized in favor of "reasonably" valid evidence needed about project performance for practical project mid-course redesign and implementation decisions.

However, this central evaluation guidance offered during the 1980s had no real force of authority behind it, was taken as little more than "suggestions" to A.I.D. missions and offices, and was frequently ignored.

By the end of the 1980s, other Agency programming and policy changes were once again affecting evaluation approaches. For example, disillusionment with basic human needs project approaches focused on isolated poverty groups led to shifts to non-project assistance modes and to emphasis on achieving policy reforms that would stimulate broad-based, market-led economic growth. Thus, evaluation methods, which for years had focused on assessing projects, now had to reorient to non-project assistance. No clear, well accepted methodology has yet emerged to rigorously assess this new form of assistance.

Also in recent years, Congress and other outside groups, as well as internal management, showed increasing concerns about the need to reorient A.I.D.'s program management and evaluation towards assessing progress in achieving strategic development goals. Isolated project "success stories" were no longer considered convincing evidence on the Hill that significant development progress was being achieved by the assistance program. A.I.D. management, especially the regional bureaus, responded by focusing more attention on developing approaches for measuring program performance and in assisting their missions with implementing program-level evaluation systems oriented around key strategic objectives and developing program performance indicators.

B. EVALUATION IN A.I.D. TODAY

This section reviews evaluation responsibilities and functions in practice in A.I.D. today at various organizational levels. Then, it reviews the distribution of resources (funding and staff) for evaluation among these levels and compares resource and staff levels devoted to evaluation as a proportion of overall program funding and manpower levels. Then, A.I.D.'s emphasis on evaluation is compared with that of other development agencies.

Distribution of Evaluation Functions and Responsibilities

A.I.D.'s evaluation system parallels the decentralized management structure of the Agency, with evaluation responsibilities divided among the central and regional bureaus, the missions, and project management levels. According to evaluation guidance, functions and responsibilities are divided as follows:

(a) Project Management's Evaluation Responsibilities: Project officers are responsible for ensuring that their project designs incorporate lessons from past experience and plan an adequate monitoring and evaluation system to meet their management information needs. They are responsible for managing, but not necessarily participating, in evaluations of their projects.

(b) Mission-Level Evaluation Responsibilities: At the mission level, responsibilities include planning and implementing portfolio-wide monitoring and evaluation systems capable of guiding project and program management decisions, reviewing the findings and recommendations of evaluations and pursuing follow-up actions, and preparing annual evaluation plans that meet mission and project management's information needs. Often an "evaluation officer" is assigned to assist the mission director with these responsibilities.

(c) Bureau-Level Evaluation Responsibilities: While evaluation's organizational structure and functions differ considerably among A.I.D./W bureaus, in general, a person or small unit in each bureau is assigned the following responsibilities: coordinating evaluation activities among missions (or offices) in order to obtain comparative data for bureau senior managers and sponsoring cross-national evaluative studies to meet these needs, preparing a bureau annual evaluation plan based on mission (or office) plans and incorporating bureau-level management information needs, establishing procedures whereby mission (or office) sponsored evaluations are sent to and reviewed by the bureaus, and providing guidance, standards, and technical assistance to missions (and offices) for monitoring and evaluation activities. In addition, other bureau staff may periodically, on an ad hoc basis, participate in evaluation activities.

(d) PPC's Center for Development Information and Evaluation Responsibilities: CDIE's evaluation functions include Agency-wide responsibilities, such as the dissemination of evaluation documents, findings, and lessons and coordination, guidance, and assistance for the decentralized evaluation system. CDIE also has special responsibility for meeting the evaluation information needs of the Agency's senior management. Specifically, responsibilities include conducting special evaluation studies relevant to (or specifically requested by) senior management; conducting cross-national evaluation studies that compare and summarize experience and lessons learned in specific program areas that will be useful for the design of similar development activities; providing guidance, standards, and technical assistance for monitoring and evaluation efforts Agency-wide; coordinating and preparing the Agency's annual evaluation planning process; disseminating information and making available on request evaluation and other project and program documents to Agency staff and external audiences.

Distribution of Resources and Staff for A.I.D. Evaluation

Based on the limited data sources available, an estimated $11 million or more of program funds was spent Agency-wide on project and program evaluation studies completed in FY1989. Of this amount, about $7.9 million was expenditure for mission-managed evaluations, and $3.1 million was for A.I.D./W sponsored evaluations, of which CDIE's share was about $1 million. In addition, several hundred thousand dollars of program funds was spent Agency-wide for evaluation systems development, including mission program performance pilots and some training workshops.

Estimates of the level of A.I.D./W direct-hire professional staff working in evaluation units (or in official evaluation positions) in FY1989 totals 18.75 person years, of which CDIE's share was 9 person years. This amounts to about 0.7% of total A.I.D. direct hire positions. Assuming an average cost of approximately $72,000 per position (including salary, office space, and other overhead), A.I.D. thus spends about $1.3 million in OE funding for A.I.D./W evaluation staff. Data on the numbers of evaluation positions in overseas missions are very spotty.

In addition, some A.I.D. staff spend small proportions of their time on evaluation activities, though evaluation may not be a regular part of their duties. A recent survey of how A.I.D.'s workforce spends its time indicates that as much as 1.7% of A.I.D.'s total workforce time in FY1989 was spent on program assessment and evaluation (4% in AID/W and 1% overseas). The figure was even higher for direct hire staff (4%). Figures 2-6 provide some details of the workforce survey, but it must be cautioned that the database has not yet been "cleaned" and may therefore be misleading.

A.I.D.'s total bilateral assistance program (including PL 480 food assistance program expenditures) totaled $7.2 billion for FY1989. Adding estimates of program-funded evaluation studies completed in FY1989 by missions and by AID/W plus the OE funding expenditures for A.I.D./W evaluation positions overhead and travel, the

Figure 2

Distribution of Work Years by Major Work Categories

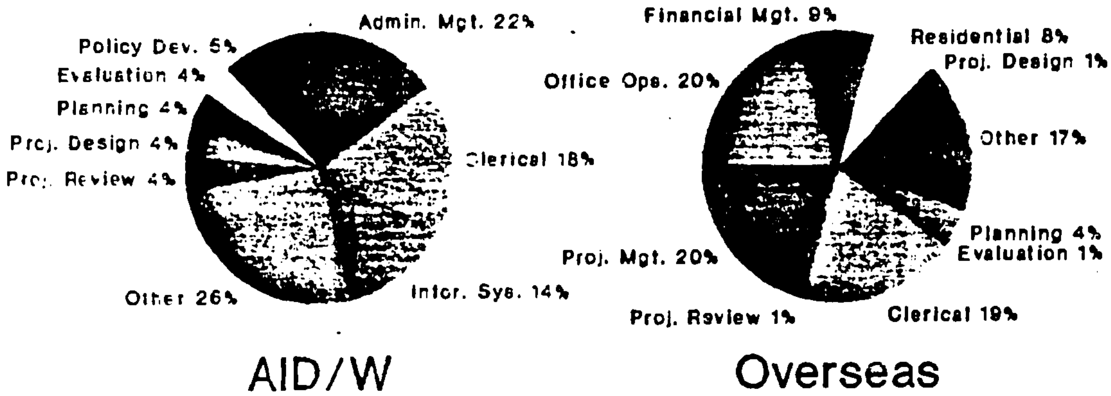


Figure 3

Distribution of Work Years by Major Work Categories U.S. Direct Hire

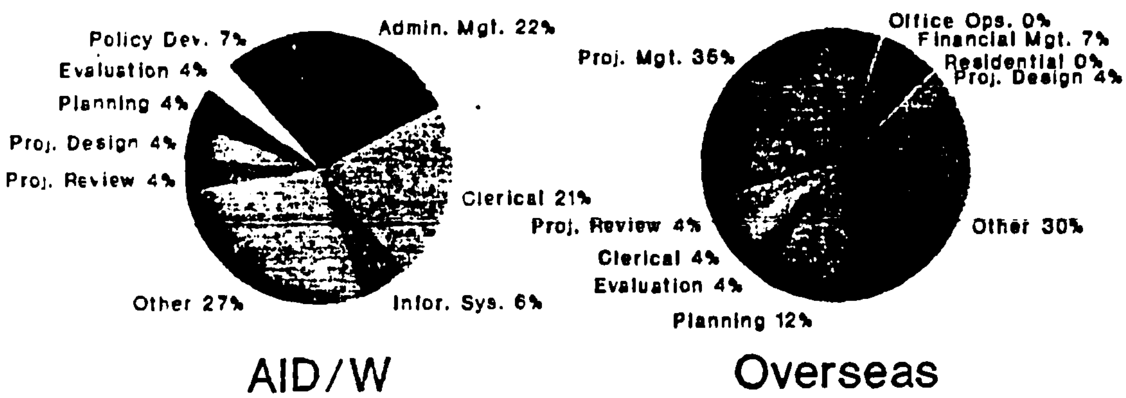
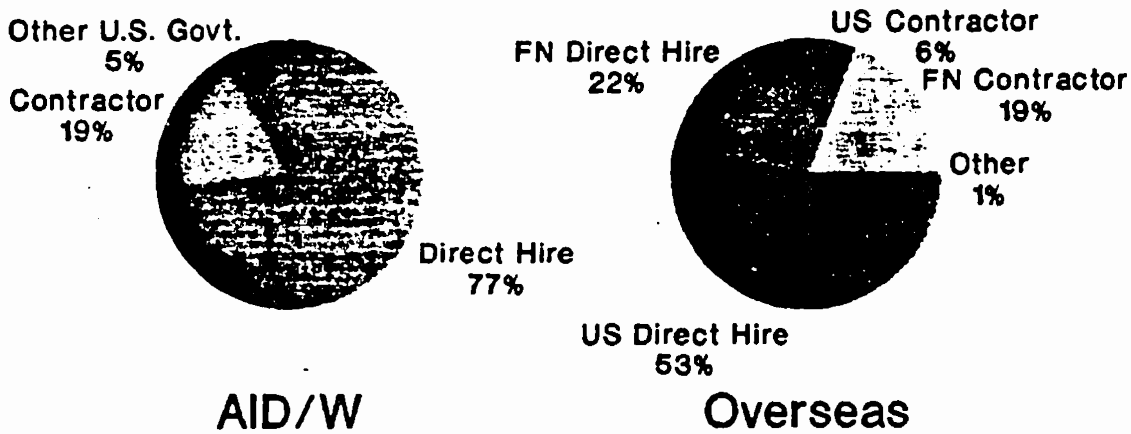


Figure 4

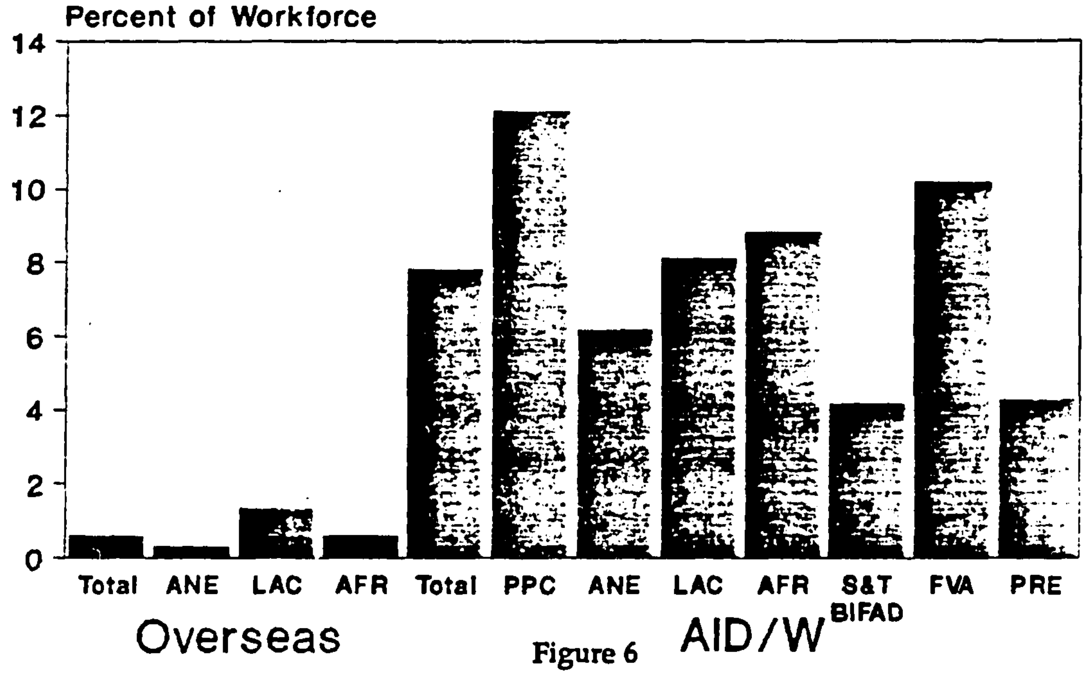
Program Assessment and Evaluation by Staff Type



Source: Preliminary ABS

Figure 5

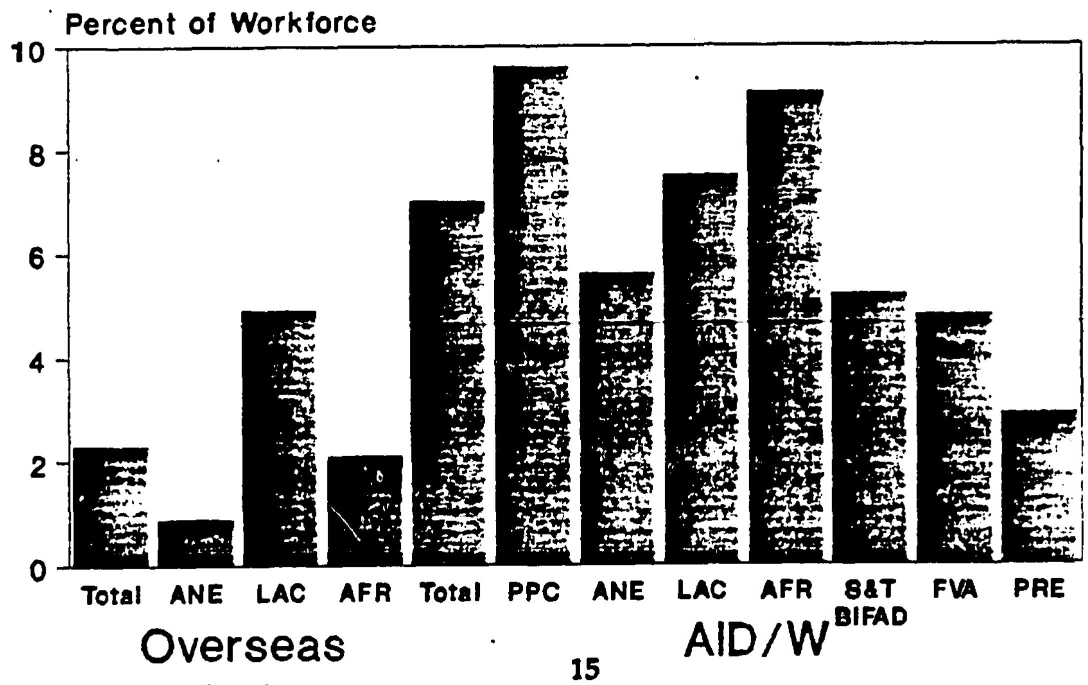
Percentage of Total A.I.D. Work Years Spent on Program Assessment and Evaluation



•Totals includes units shown

Figure 6

Percentage of Total A.I.D. US Direct Hire Work Years Spent on Program Assessment and Evaluation



•Totals includes units shown

Agency spent only a small fraction of 1% (less than 0.2 %) on evaluation.

Comparisons of A.I.D./W staff in evaluation positions (18.75 person years) with those working in audit positions in AID/W (107 person years), indicates that there are about five times as many auditors as there are evaluators in AID/W.

Comparisons with other Donor Agencies' Evaluation Systems

A 1987 study of multilateral donor agencies' evaluation systems provides some useful benchmarks or indicators for comparing A.I.D.'s evaluation system. This study by the Overseas Development Administration examined expenditures and staff for what they called "independent" evaluation units only; that is, excluding "self-assessments" sponsored by project or program management. This would be roughly equivalent in concept to A.I.D./W centrally sponsored evaluations, thus excluding the missions' evaluation efforts. As Figures 7-8 indicate, A.I.D. spends a relatively small percentage of its total program expenditures on "independent" evaluations (0.07%) compared with most multilateral development agencies. Similarly, the percentage of A.I.D. professional direct-hire staff working for independent (i.e., A.I.D./W) evaluation units (0.7%) is low compared with many of the multilateral agencies such as the World Bank (1.2%), the Asian Development Bank (2.7%), IFAD (4.1%), or UNFPA (3.9%).

Comparisons with other U.S. Agency Evaluation Systems

Unfortunately, no similar study has been undertaken recently that gathers similar indicators on evaluation expenditures and staff of U.S. domestic agencies. However, a 1987 GAO report on federal evaluation indicates that some federal agencies have instituted standards for evaluation spending through the mechanism of set-asides. For example, the National Institutes of Health has a set-aside policy in which 1% of the funds allocated for individual projects can be used for evaluation. In addition, the central evaluation unit of NIH has a yearly budget for evaluation funds to which the institutes within NIH may apply for additional evaluation resources. The Public Health Service has a Congressional authorization which allows up to 1% of dollars appropriated for public health service programs to be used for evaluation. This set-aside is delegated to the Assistant Secretary for Planning and Evaluation for implementation. A similar legislative set-aside of up to 1% for evaluation exists for the USDA's Food and Nutrition Service's programs.

Figure 7

Percentage of Donor Program Expenditure Spent on Independent Evaluations

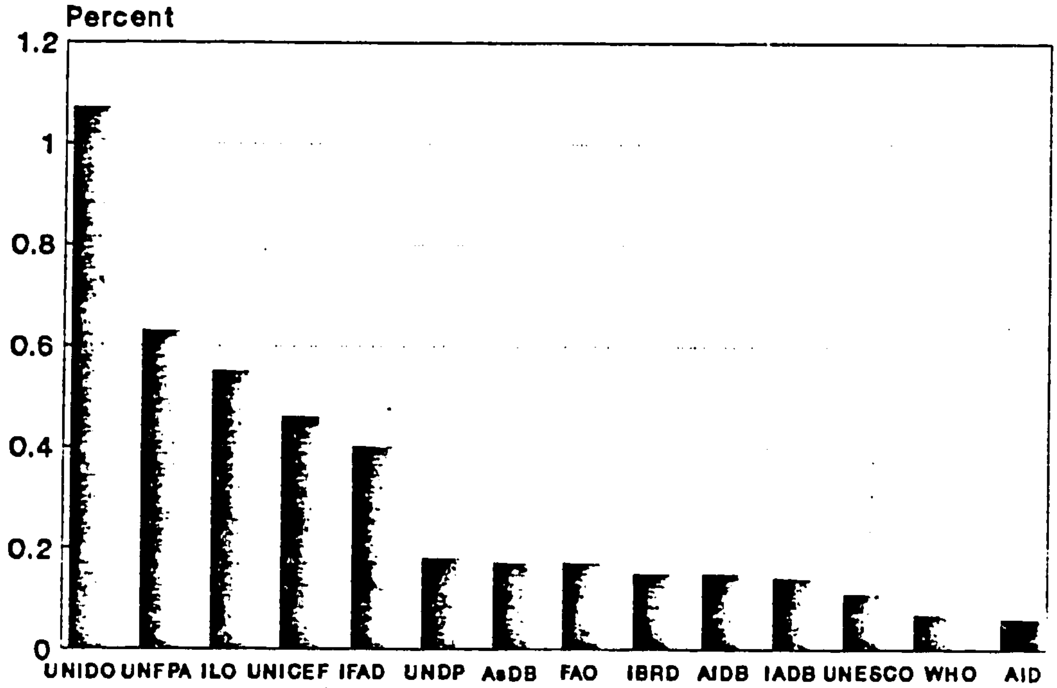
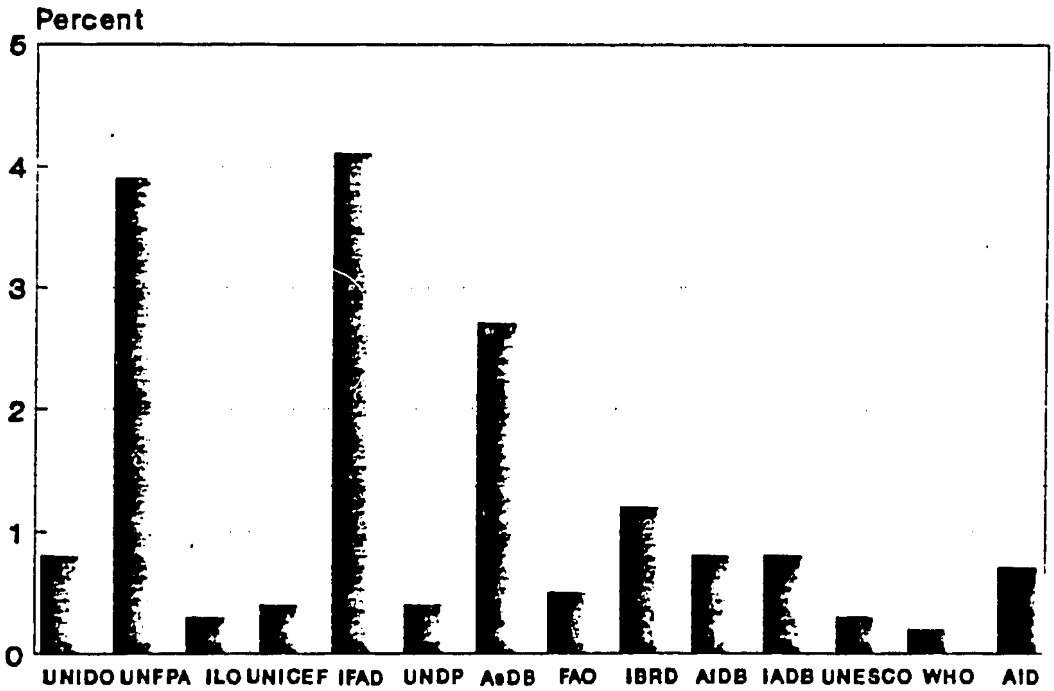


Figure 8

Percentage of Donor Professional Staff Working in Independent Evaluation Units



IV. ASSESSMENT OF EVALUATION IN A.I.D. TODAY: PROGRESS AND PROBLEMS

This section reviews the A.I.D. evaluation system's progress achieved, the significant remaining problems, and their probable causes. The critiques made are based on perspectives of what potential purposes, audiences, and information needs the evaluation system should be possibly serving (referring to the theoretical framework developed in earlier sections of this paper). The assessment reviews evaluation efforts at different organizational levels, starting with project level, then mission level, and finally A.I.D./W levels.

A. EVALUATION AT VARIOUS ORGANIZATIONAL LEVELS

Project-Level Evaluations

A.I.D.'s decentralized evaluation system produces about 250 evaluation reports sponsored by missions annually. The vast majority of these studies focus on individual interventions. They make up about 85 percent of all Agency evaluations, with the remainder being regional or central bureau sponsored evaluations.

Having such a decentralized evaluation system makes sense given the decentralized organizational and decision-making structure of the Agency. Experience shows that utilization is promoted by placing evaluation systems and units organizationally close to the management structure it is intended to serve. Decentralization of evaluation is likely to enhance its use in the field, for example, in improving project implementation and redesign efforts.

However, decentralization of the evaluation system has created some problems, especially its ability to respond to some of A.I.D. senior management's or Congress's evaluation information needs. The system in general has not produced evaluations with a common scope, focus, methods, or standard indicators that would allow comparative analysis or aggregation of results across programs or countries. Moreover, since project evaluation is basically optional, it is impossible to get complete coverage of performance in any program area Agency-wide. It is difficult to respond to calls for program "accountability" or "performance based budgeting" in such circumstances. There are some exceptions to this, for example, in the more centrally run programs of family planning and child survival, where more standardized and quantitative evaluation methods have been more regularly applied. Also, the promising new initiatives in developing mission-level program performance management and evaluation systems may eventually help the decentralized evaluation system to develop standardized indicators in core program areas and to respond more appropriately to senior management's and external audiences' needs. Another thorny issue that would need to be resolved if mission evaluation reporting were to be seriously used for A.I.D./W resource allocation decisions or for judging accountability for performance would be how to avoid biases in findings and reporting by parties with obvious vested interests.

Even in terms of their abilities to serve immediate project-level operational decision needs, the decentralized evaluation system has some significant shortcomings. A recent central review of some 300 mission evaluation reports completed in FY 1987-88 found that while they were serving important monitoring functions, they typically were not answering higher level questions about performance and results adequately. Central guidance, prepared to strengthen evaluation planning, methods, and utilization, was not being regularly followed by the field. Empirical data on project performance and results were most often unavailable; monitoring and evaluation systems within projects were missing or inadequate, and neither formal surveys nor more innovative low-cost qualitative evaluation methods were being systematically used. Moreover, not all evaluation report recommendations were found to be useful or actionable by the project or mission management sponsoring them. Collaborative evaluations as encouraged by the guidance, although promising, were not yet the rule. (See appendix B for details.)

Other surveys have corroborated some of these findings. For example, a review of 60 project papers prepared in 1989, about two years after the new Evaluation Handbook was circulated, assessed whether they were incorporating plans for establishing ongoing monitoring and evaluation systems within projects in compliance with the guidance. Results indicated that only roughly one-quarter of the project papers included reasonably thought-out monitoring and evaluation system plans. Another review examined 75 program and project audits prepared by the Office of the Inspector General during FY1987-88 in terms of their comments about the adequacy of monitoring and evaluation. Over one-half cited generally ineffective monitoring and little or no effort to evaluate performance. About one-quarter of the audits discussed specific monitoring and evaluation inadequacies and failures at the purpose/goal level, while only a small fraction (4%) made positive references to monitoring and evaluation improvements or capabilities.

Mission-Level Evaluation

While undoubtedly some missions had model management information systems that systematically applied evaluation experiences and lessons to mission-level strategic program planning and allocation decisions, this was probably not typical of many missions in the 1980s. Evaluation findings, recommendations, and uses focused on the level of individual interventions rather than on program- or mission-level decisions, not to mention potential A.I.D./W needs.

Over the last decade, CDIE and regional bureau evaluation units provided highly selective technical assistance to missions to attempt to improve project-level monitoring and evaluation. However, given the small evaluation staffs of the A.I.D./W bureaus, only a small fraction of mission projects received any sort of technical assistance in planning evaluation systems or in undertaking evaluations. Another effort to improve the quality of decentralized evaluations was to prepare a special series of centrally sponsored guidance reports on evaluation methods during the 1980s. However, by the end of the 1980s, it was becoming increasingly clear that the guidance reports alone were not making a significant enough difference in the decentralized evaluation system, and that a more proactive effort would be needed to visibly improve mission evaluation systems.

By the end of the 1980s, pressure from Congressional and other outside interest groups mounted on the Agency to reorient program management and evaluation towards concern and accountability for development results. A.I.D. management, especially in the regional bureaus, responded by focusing more attention on developing approaches for measuring program performance and in assisting their missions with implementing program-level evaluation systems oriented around key strategic development objectives. The focus on performance indicators at the program level, as opposed to project level measures, paralleled a broader Agency shift in recent years from an exclusively project approach focused on pockets of poverty, to concern with broader strategic or "program" objectives aimed at making a significant difference for overall host country development. Programs often encompass a variety of coordinated assistance modes (including for example, project, non-project, and policy dialogue approaches), all contributing to the same objective.

Searching for ways to strengthen A.I.D.'s decentralized evaluation system, the regional bureaus and CDIE about a year ago began to examine ways to evaluate program performance, develop mission program performance indicators, and more strategically manage programs for development results. CDIE began to work with the bureaus to help coordinate, standardize, and guide these efforts. Seven missions were selected out of dozens of interested missions to serve as "pilots" aimed at developing model program performance management and evaluation systems. Related efforts were undertaken to strengthen mission evaluation systems in another dozen missions.

The program performance pilot efforts involve (1) initial strategic planning and information needs assessments to clarify mission program strategies, objectives, indicators, and information/data needs; (2) design and implemention of appropriate program performance monitoring, reporting, and evaluation systems for the mission; and (3) assistance in applying program performance information in mission management decisions and as a basis for upward reporting to regional bureaus, senior management, and the Congress.

An initial review of these pilot efforts demonstrates that they are important initiatives for strengthening the decentralized evaluation system. They are potentially useful to all Agency management levels. If sustained and replicated, they hold promise for making a number of improvements:

- The initial strategic planning stage is helping missions to think through, focus, and clarify what their strategic goals are, how their various interventions are linked to these programmatic objectives, and how performance can be measured.

- The focus of evaluation efforts in missions will be more clearly on measuring results or impacts, rather than just on monitoring concerns.

- Appropriate systems for gathering more rigorous, empirical data on impacts are being put in place, ensuring the ongoing measurement of program and project performance.

- The pilots appear to be elevating the position of evaluation systems and use of findings to a more central status in mission management's attention. The pilots are demonstrating to mission management that performance information can be directly useful for making operational decisions about their programs, and also for "reporting up" to A.I.D./W about the effectiveness of their programs.

- The pilots are revealing some interesting points about A.I.D.'s assistance program. For example, it appears that the diversity in Agency objectives and in program approaches is not as great as many had perceived. Most missions focus on about three to six strategic objectives, while for the Agency as a whole, there seem to be about a dozen "core" program elements that keep appearing again and again. This commonality in goals and consistency in intervention approaches across countries means that the development of more comparative, standardized performance indicators around clearly delineated program objectives may be possible in the not too distant future. Eventually, a system of comparable, empirically based indicators of missions' program performance could become of use to A.I.D./W senior management for comparatively assessing country and program performance Agency-wide.

Central-Level Evaluation

Beginning in the early 1980s, the central evaluation office (CDIE and its predecessor), and later the regional bureaus, began to undertake a series of cross-country, comparative evaluations focused on investigating what the development results and experiences of A.I.D. interventions have been within priority program areas. Unlike most mission evaluations which focused directly on improving project implementation processes, these centrally sponsored evaluations generally had broader objectives — analyzing, comparing, and synthesizing findings from a series of development interventions to provide a basis for future Agency-wide program and policy guidance. By focusing "above" the individual project level, these evaluations helped to meet senior management's needs for synthesized and comparative information as a basis for influencing A.I.D. policy and program formulation.

CDIE, often in collaboration with other regional and central bureau offices, has published over 200 evaluation studies over the last decade. These documents provide considerable knowledge and lessons about the Agency's experience in numerous program areas, e.g. roads, electricity, potable water, irrigation, agricultural research, agricultural credit and other services, agricultural universities, integrated rural development, microenterprise, policy reform programs, and health. Also, some studies focus on "cross-cutting" policy themes, such as impacts on achieving women-in-development objectives, or how to improve program management and sustainability.

The methodologies used in these studies have varied, depending on factors such as time and resources available and level of interest. Some approaches have relied on short-term fieldwork in a number of countries, using a common scope of work to assess and compare interventions, while other studies have relied on "desk studies" involving a review and synthesis of existing evaluations previously performed by missions. In general, however, these methods have been low-cost, qualitative, rapid appraisal approaches. Rarely did they attempt longer term, more rigorous statistical approaches that could accurately measure impacts or cost-effectiveness. Indeed, limited resources, staff, and technical capacities in the 1980s prevented attempting such rigorous methods. Moreover, there was no real expressed demand for such studies coming from senior management or from Congress until the end of the decade.

Two assessments of CDIE's evaluation studies program in the late 1980s argued that greater efforts were needed to make the evaluations more relevant to senior management and useful in the Agency's policy, program, and budget decision-making processes. They suggested more involvement by senior management in setting the priorities for the CDIE evaluation agenda and greater CDIE networking and coordination among offices and with the field. They also argued that the evaluation studies were not being regularly read nor were lessons frequently enough used by A.I.D. management, and that more attention needed to go into report quality, timeliness, presentation techniques, and "marketing" of findings.

CDIE's response has been, in part, a revitalization of the Agency's traditional impact evaluation series, first initiated in 1979 by the A.I.D. Administrator. The recent impact evaluations have focused specifically on A.I.D. senior management as the primary audience. CDIE has worked in close collaboration with the regional and technical bureaus in planning these evaluations, in some cases using joint funding. The methodology employed by these impact evaluations has been a "rapid appraisal evaluation" approach, not to be confused with the conventional statistical "impact evaluation" methodology discussed in Appendix A.

The CDIE approach is based on sending a small multidisciplinary team to the field for a short three-week period for intensive interviewing, observation, and information gathering about project results and factors affecting results. While teams are expected to assess and reanalyze existing quantitative data, they are not expected to undertake any new statistical surveys; however, they are encouraged to use systematic rapid appraisal techniques. Currently, the series are focused on policy reform, child survival, and family planning programs. Emphasis has been placed on preparing eye-catching, highly readable and brief reports with operationally useful lessons. Senior managers have been recruited as leaders of the direct-hire dominated impact evaluation teams to help ensure relevance of findings to the senior management audience. Direct participation by A.I.D. staff in the evaluation process was also seen as an important approach for helping to refocus A.I.D. management's attention on development results.

Several factors can be identified that may limit the approach taken by these impact evaluation efforts. One such factor is the lack of authority to evaluate any interventions regardless of missions' concurrence. This may well lead to an immediate bias in the sample of projects selected to study, because missions may be less likely to give their approval if they consider the project a failure. Other considerations relate to the quality and independence, or objectivity, of the evaluation teams. Recent emphasis in the CDIE impact evaluation series on using A.I.D. direct-hire staff exclusively, and on using different people for each field study, may have resulted in somewhat ad hoc teams, with limited technical or methodological capabilities, and with varying or unstructured approaches to the evaluation. Despite team-planning workshops and common scopes of work, differently composed teams tend to approach the evaluation task differently. Use of direct-hire staff, especially in cases with past associations with the intervention being evaluated, may cast doubt on their objectivity. It should be noted, however, that using outside contractors does not guarantee objectivity, especially if the firms or individuals doing the evaluations do most of their business in related design and implementation work for the Agency. Also, use of busy A.I.D. senior managers to lead the teams has meant that the efforts are often squeezed into too short a time frame in order to accommodate their schedules, and trying to get revisions or final reports from them once they are back in A.I.D./W is often very difficult.

CDIE has also continued to undertake other special evaluation studies, using different methodologies and levels of effort. Some, such as a current effort in evaluating trade and investment promotion projects and a proposed study of narcotics control assistance programs, will be fairly substantive, rigorous studies involving some primary data collection and analysis by contractors in several countries. However, CDIE has not had the mandate nor the resources in the past to do much long-term scientifically based survey research on program impacts.

At the other end of the scale, CDIE also continues to do low-cost, quick-turnaround "desk studies" that rely on synthesizing information on specific topics from existing mission evaluations available in A.I.D./W, rather than going to the field. While synthesis efforts are often constrained by the variability in the quality and lack of comparability of existing evaluation reports, they nevertheless have proven useful in extracting lessons for future Agency program design guidance.

In addition, several new series of very short "applications" products have been developed recently to help summarize and market Agency evaluation study findings for specific A.I.D. audiences. These include the series of Evaluation Highlights and Innovative Development Approaches reports, the two-page Bulletins for Senior Managers, and the Evaluation Newsletter.

CDIE also provides a variety of Agency-wide services in support of the evaluation function. Over the years this support has included responding to Agency staff requests for evaluation documents or summaries of experience; developing evaluation training courses, methods workshops, and methods guidance papers; providing technical assistance, and, until recently, coordinating the evaluation planning process.

The annual evaluation planning process never worked as planned in the guidance. Lacking in management attention and participation, especially at the higher A.I.D./W levels, it soon became a "paper" exercise focused on tracking whether planned mission evaluations were actually done or not. There was little payoff in terms of management establishing priorities for evaluation or coordinating evaluation needs among different management levels, offices, or missions. Due to lack of interest (and staffing cuts), CDIE stopped preparing annual Agency Evaluation Plans in the mid-1980s.

CDIE and other A.I.D./W bureau evaluation efforts have not gone very far in meeting some of the senior management needs discussed earlier in this report--particularly the need for more long-term, rigorous, statistical, and comparative evaluations that could serve as a basis for performance-based budget allocation decisions. This has not been an explicit mandate in the past, nor have adequate resources for such efforts been available, except on an occasional, experimental basis. A.I.D./W sponsored evaluations have been more useful for influencing Agency policy guidance and providing lessons for future project design efforts.

The growing demands by Congress and other oversight agencies for more comprehensive and convincing evidence on the development results of A.I.D.'s programs also is not being adequately addressed by central and regional bureau evaluation offices. While promising program performance pilot efforts are underway in several missions, for these to be truly useful for addressing calls for accountability, much more will need to be done to expand these efforts and make them more comparable and free from potential bias.

B. SOME UNDERLYING REASONS FOR CONTINUING PROBLEMS

1. Low Management Priority for Evaluation: A fundamental reason for continuing inadequacies in evaluation in A.I.D. is that much of the

Agency's management appears to assign evaluations a relatively low priority. Some possible reasons may include the following:

- Lack of signals and actions in recent years from the Administrator's office to the rest of the Agency indicating the importance of evaluation and management's accountability for development results.
- A heavy paperwork burden on a declining number of A.I.D. staff that emphasizes extensive planning and design requirements (despite redelegations of authority to the field), voluminous reporting requirements to Congress, complicated contracting and procurement regulations, etc. leaving little time for thinking about achieving development results or for tracking performance.
- An annual budget cycle that encourages priority to be placed on program planning and on obligation of money as the major Agency focus for action, rather than on the achievement of development results.
- The common perspective in A.I.D. that funding allocation and programming decisions will be made based on foreign policy objectives and on earmarking, rather than on development performance.
- An Agency personnel system that offers little career incentive for managers to focus on development results or evaluation work.

2. Lack of a Clear Mandate for Central Evaluation: For many years, there has been no clear leadership from the Administrator's office to guide the central evaluation office's work agenda. Nor has a clear mandate been given to this office to serve as an independent, autonomous judge of the Agency's program performance and development results. The existing structure requires the central evaluation office to operate without much authority, undermining its effort to conduct rigorous, objective analyses of A.I.D. activities or to review the quality standards of evaluation work being performed by other offices and missions.

3. Emphasis on Mission Management Needs in Evaluation: The decentralization of program approval authority to missions took place without developing a system for mission management accountability for achieving and reporting results to A.I.D./W senior management. There is now a growing interest in developing such reporting systems. Furthermore, coordination of evaluation information needs among management levels has been lacking. As a result, the evaluation system has tended to focus until recently on project and mission needs, leaving senior management and Congressional needs largely unmet.

4. Limited Evaluation Methods: Senior management and Congressional information needs in particular have not been adequately met in part due to lack of appropriately rigorous and objective methods. These in turn have been limited by lack of funds, time, staff capabilities, methodological difficulties, and lack of a clear mandate to undertake rigorous, comparative, centrally sponsored evaluation studies.

5. Limited Evaluation Staff and Resources: Finally, one cannot ignore the factor of the relative lack of staff and funding devoted to the evaluation function compared with other functions. For example, staff working primarily on evaluation numbers fewer than 20 in A.I.D./W, compared with over five times this number of auditor positions. Training courses for A.I.D. staff focused on evaluation methods have diminished in frequency. Total funding expenditures on evaluation studies may be as low as $11 million per year Agency-wide; thus a small fraction of one percent of the U.S. assistance program funds are spent on evaluation of the Agency's programs.

During the 1980s, the trends in staff and resources devoted to the central evaluation function has been steadily deteriorating. For example, CDIE's evaluation staff fell from about 28 direct-hire positions at the time of its establishment to 9 today, while CDIE's annual budget levels for evaluation fell by half from about $2 million to $1 million.

V. SUGGESTIONS FOR STRENGTHENING EVALUATION IN A.I.D.

These observations on the history, progress, and problems of A.I.D.'s evaluation system, when assessed in the context of the purposes and audiences it perhaps should be serving, lead to a number of suggestions for how to improve evaluation in A.I.D. First, we look at ways to improve the quality and the relevance of evaluation for these key purposes and audiences, and then at ways to remove constraints. It should be noted that some of these suggested actions could be undertaken quickly once decisions to do so have been made, while others would require considerably more time or even substantial organizational changes and resource shifts to implement. Also, some of the suggested actions will require the Administrator's personal leadership to be effective, while other actions could be implemented at the Bureau or lower levels.

A. IMPROVING THE QUALITY AND RELEVANCE OF A.I.D. EVALUATIONS

How can the Agency's evaluation methods, practices, and approaches be strengthened (a) to improve the integrity and intellectual rigor of A.I.D.evaluations, and (b) to better meet the needs of project and mission management, senior A.I.D./W management, and external Congressional and constituency groups? Overall, there is a clear need to place more emphasis on intellectually independent, rigorous, and empirically-based methods that focus on higher level program performance and development results.

1. Project Level Evaluation Systems: In order to improve evaluation methods, practices and approaches that would better serve project management's operational decision needs, and facilitate their utilization, the following steps might be taken:

(a) Establish incentives and disincentives to encourage higher quality evaluation work by field missions. For example, institute some AID/W oversight of mission evaluation performance, such as tasking CDIE or other AID/W offices (or even the IG) to selectively review the scopes of work for major mission evaluations, or assess the adequacy of monitoring and evaluation system plans in important project design documents. Another approach would be to randomly audit or review several missions in terms of the adequacy of their evaluation systems and utilization of evaluation lessons and recommendations.

(b) Reinvigorate AID/W-funded training programs in evaluation methods and procedures for A.I.D. mission project officers and host country counterparts from both the public and private sector. Such training programs could focus on program-level evaluation systems to reinforce the Agency's new strategic direction and on the use of both quantitative and qualitative evaluation techniques.

(c) Encourage more collaborative evaluations and development of local institutional capacity to evaluate, not just in the public sector but also in private sector research firms and local universities.

(d) Provide A.I.D. missions with ready access, through separate evaluation-only IQC's

or other contracting vehicles, to a pool of disinterested, expert evaluators. Such contracting mechanisms could help diminish conflicts of interest many contractors now face when doing A.I.D. program design and implementation work as well as evaluation.

2. Mission Management Evaluation Systems: To strengthen the quality of mission-level evaluation systems and practices and to encourage their use by mission management in strategic program planning processes, several actions could be taken:

(a) Accelerate and expand the development and implementation of the promising program performance and evaluation systems now being tested in some twenty missions. Promote central and regional bureau technical assistance to missions in these efforts and develop general guidance in setting up such systems.

(b) Require that CDSS's and other country strategic planning and reporting documents include more quantitative data and analysis of program performance indicators, gathered by such evaluation systems.

(c) Encourage participation by the host country governments in program-level evaluation planning to build a constituency within recipient countries for more strategic use of evaluation information; this corresponds well with A.I.D.'s priority on policy dialogue and reform.

3. Central Evaluation Systems: To strengthen the capacity of central (A.I.D./W) evaluation units to serve senior management and external oversight and constituency groups, evaluation methods and approaches should:

(a) Continue to strengthen centrally sponsored "rapid appraisal" field-based evaluations by (i) developing more systematic research methodologies; (ii) ensuring that team composition reflects mix of in-house and outside expertise and avoids conflicts of interest; (iii) raising technical and/or methodological expertise of teams; and (iv) empowering the efforts with greater AID/W authority to select appropriate intervention sites for evaluations once the agenda has been set by senior management.

(b) Initiate and test "state of the art" methodologies for more rigorous, longer term, comparative studies in selected program areas. Such studies serve to guide senior management's resource allocation decisions by providing empirical evidence of the relative performance, impact, efficiency, (i.e., cost-benefit and cost-effectiveness) and sustainability of alternative intervention approaches.

(c) Explore other, lower cost options for generating findings on longer term development results, such as offering matching funds for graduate-level field research by both U.S. and LDC scholars on the impact of foreign assistance in selected countries or programs.

(d) Building on the mission-level program performance evaluation systems, establish an Agency-wide information system for reporting on program performance to A.I.D. senior management and to Congress by (i) institutionalizing the capacity to collect program performance data Agency-wide, and (ii) developing standard Agency-wide program performance indicators in core program areas to facilitate cross-country assessments.

B. REMOVAL OF CONSTRAINTS

Setting an agenda for improving the quality and relevance of evaluation methods, practices, and approaches is only part of what needs to be done to strengthen evaluation in A.I.D. The other aspect involves removing or mitigating a host of other constraints, including limited management interest in or time for evaluation; practical resource, staff, and time constraints; disincentives in the personnel system; lack of a clear central evaluation mandate and adequate authority. The following discussion outlines steps that could be taken to help remove such constraints.

1. Greater management, especially senior management, involvement in the evaluation process. Perhaps most important for stimulating change and improvements in the Agency's evaluation system is the need for clear signals from the Administrator and other senior officials that they consider the evaluation function to be important. Specific actions that might be taken include the following:

(a) **More direct involvement by the Administrator and senior management in setting the agenda for evaluation.** Decision-makers need to communicate more with evaluation planners at all management levels to ensure that their priority information needs are met by the evaluation system. One possibility for accomplishing this might be reinstituting the annual evaluation planning process, but this time ensuring management participation at all levels, as was originally envisioned. The process might culminate in a collaboratively developed Annual Report to the Administrator on the status of evaluation planning and progress in A.I.D. Setting aside a few sessions of the Administrator's Review for discussion of the Agency's evaluation agenda might also serve this purpose. Such meetings should include discussions of what topics and program areas are of priority interest to senior management and should be included in centrally sponsored evaluation studies.

(b) **Encourage senior managers to make time available to take an active role in evaluation work.** For example, the Administrator's visible support for releasing senior managers from their everyday responsibilities to lead short-term impact evaluation field studies would send a signal about the priority placed on evaluation, and would serve to reorient management concerns from routine program administration and implementation to a focus on whether we're achieving our longer term development objectives.

(c) **Encourage the Agency's management to utilize evaluation findings and apply lessons from experience to future program and policy decisions.** Steps that might be taken include greater participation by the Administrator and senior managers in conferences, workshops, and debriefings on major program evaluation findings. Another step might be to encourage greater participation by evaluators in management meetings where key program and policy decisions are to be made, so that they may benefit from relevant evaluation findings and lessons.

(d) **Reduce other demands on management time and priorities that keep them from focusing on program performance and results.** Some of the Agency's paperwork burden is self-imposed while others are Congressional or regulatory requirements (e.g., obligation deadlines, contractual and procurement rules, earmarking, Congressional reporting requirements on program plans). Over time A.I.D. senior management could do something about shifting internal work priorities, and perhaps through negotiations with Congress might achieve a "trade" whereby the Agency could gain greater upfront programming flexibility in exchange for more rigorous and convincing evidence that A.I.D. has a results-oriented, performance-based management system, and that development assistance programs are achieving their goals. In effect, such a deal was reached between A.I.D. and the Congress in the Africa Development Fund, and the Hamilton Task Force last year was advocating that these sorts of changes become part of the aid legislation in general.

2. **Place greater emphasis on senior management's evaluation information needs.** While maintaining a decentralized evaluation structure is appropriate for the A.I.D. organizational structure, emphasis within the system probably needs to be reoriented more towards senior management and Congressional requirements to redress past shortcomings. The decentralized evaluation system needs to be "harnessed" to serve A.I.D./W needs as well as mission and project level needs, and centrally sponsored evaluation studies promoted that are directly relevant to senior management.

3. **Provide a clearer mandate for the central evaluation office, with authority to independently select and evaluate projects and programs based on objective criteria, without requiring excessive concurrences by other parties.** The Administrator's direct involvement in setting the agenda and topics for the central evaluation office's evaluation studies should by itself go a long way in establishing such authority and independence vis-a-vis the rest of the Agency. In addition, the purposes and audiences that the central office serves need clarification. For example, if centrally sponsored studies are ex-

pected to address program resource allocation decisions or to provide accountability reporting to Congress this needs explicit recognition. Expectations concerning methodologies also require discussion, especially if use of cost-benefit, cost-effectiveness, conventional impact evaluation research designs, or other rigorous quantitative approaches are indicated.

4. Implement personnel and career incentives that will encourage higher quality evaluation work and broader use of evaluation by managers.

(a) The Agency needs to broaden its definition of "accountability" to include management accountability for development results. This will involve a dramatic and possibly difficult shift in focus from obligating and routinely implementing programs to a focus on achieving results from those programs.

(b) The current system of personnel performance appraisal ratings might be reviewed to see if more emphasis could be placed upon incentives to improve evaluation efforts and the utilization of findings. While managers cannot be held directly responsible for the achievement of development results of projects and programs, they can be expected to ensure that their management reflects a focus on results. Personnel ratings could reflect the extent to which managers display a concern for measuring and evaluating performance and undertake corrective actions based on evaluation findings.

(c) Career incentives could be provided for staff to move into evaluation positions. Currently, A.I.D. managers view program planning and design jobs as better stepping stones for advancement than evaluation jobs in an Agency that appears to value "moving money" more than achieving or analyzing development results. Experience in evaluation work should be given greater weight in decision to promote individuals into EPAP and the Senior Foreign and Executive Services.

5. Improve the adequacy of Agency resources and staff devoted to the evaluation function and reconsider the adequacy of its organizational placement.

(a) A review should be undertaken of the adequacy of the Agency's resources and staff positions allocated to evaluation. Only a small fraction of 1% (less than 0.2%) of A.I.D.'s bilateral assistance program funds are spent on evaluation. Less than 1% (0.7%) of A.I.D.'s direct-hire positions work primarily on evaluation functions. Other donor agencies allocate a larger share of their resources for evaluation than does A.I.D. Some U.S. federal agencies have set-asides for the evaluation of major domestic programs of up to 1% of total program costs. Clearly, if broader evaluation functions and responsibilities, and new initiatives involving more rigorous methods, are to be undertaken in A.I.D., there will be obvious implications for increasing resources in support of these new challenges.

(b) The review should address whether the current allocation of resources and organizational placement of evaluation functions is optimal. Evaluation units appear to function best when they are placed organizationally close to the management levels they are intended to serve. This implies for A.I.D. the continuation of a multi-level system of evaluation units paralleling the broader management organization of the Agency. However, it is unclear whether the current mix of central versus regional and field level evaluation resources is the most adequate, or if the current organizational placement, independence, and authority of the central evaluation office is optimal given its current and future functions. These matters need to be debated and resolved.

(c) A core staff of independent evaluators with technical and methodological expertise will need to be recruited, contracted, or obtained through training. To some extent, constraints will not be only a question of hiring existing expertise, but of efforts to develop appropriate methodologies that do not yet exist or that are in their infancy, such as evaluation of non-project assistance modes, program-level assessments, or sustainability analyses, to name a few.

APPENDIX A

(a) Issues in Implementation Monitoring

In A.I.D., no clear distinction exists between monitoring and evaluation. Here, monitoring is defined as studies investigating the timely implementation of project activities (i.e., inputs, outputs) in accordance to design targets and schedules, and with assessing implementation problems, such as contractual, procurement, or funding delays, or management and staffing constraints. While the monitoring function is essential, it does not focus on the achievement of development results of the project activities, even on more immediate effects.

(b) Issues in Effectiveness Evaluation

This level of evaluation focuses on the effectiveness or performance of an intervention in meeting its purpose. It typically examines the extent to which an intervention's outputs, products, technologies, or services are being used effectively by the intended target population. Data collection requirements for such analyses are relatively simple and can often be obtained from project client records, or from "mini" surveys with samples as small as 100 respondents. While beginning to get at development results, these indicators tend to be closely related to program performance and intermediate outcomes, but stop short of assessing longer term, goal-level impacts.

Many evaluations include an analysis of the factors that influenced a project's successful or unsuccessful performance. This analysis of factors is important if operationally useful lessons are to be drawn from the evaluation. These factors may include aspects internal to the project and thus within the control of project management, such as organizational and management approaches, the distribution system, the appropriateness of the technology or services being promoted, the extent of community participation, etc. Other factors influencing project performance may be external to the project and thus more difficult for project managers to influence, such as host government commitment, favorable policy environment, sociocultural setting, political stability, the weather, and other environmental or contextual factors.

(c) Issues in Impact Evaluation

Impact evaluation attempts to measure the degree and direction of ultimate, or goal-level changes resulting from assistance efforts--most typically changes in the incomes, welfare, or behavior of the beneficiary population. For example, in the case of child survival projects, such assessments would measure changes in infant and child mortality caused by a particular intervention. Such assessments require repeated statistical surveys tracking indicators of the impact changes occurring over time, but also require analyses that can legitimately link or attribute at least some of the changes to a particular intervention. Often this is attempted by comparing changes in project areas with control groups or by using statistical controls. Proving attribution to specific interventions has proven to be difficult in all situations, but is especially so in developing country contexts where a wide range of external factors unrelated to project assistance may affect outcomes, and where practical constraints to data collection and analysis can be overwhelming.

These conventional methodological strategies for measuring net impacts attributable to specific interventions — quasi-experimental designs and statistical controls —have been used successfully in some program areas, such as in family planning and child survival, where goals are clear, where standard indicators are widely accepted, and where adequate resources have been devoted to the impact evaluation efforts. However, these methods are rarely used for much of A.I.D.'s program portfolio. Some of the problems typically encountered in their use include the following:

- **High Costs:** Requirements for rigorous, multiple rounds of sample surveys have made these methods very costly, often running into the hundreds of thousands of dollars.

- **Lengthy:** The methods require a long time frame to complete, from project start through to project completion or beyond, before evaluation results are available. Because of this, they have been of little use to the project manager concerned with improving implementation, and have also been hard to fund and complete within the context of a project. For these reasons, many of these efforts were never completed.

- **Dependence on Specialized Skills:** The approaches require rigorous statistical and data collection skills often beyond the capabilities of host country and A.I.D. staff. Yet reliance on external "experts" often resulted in evaluation issues and findings being oriented away from management needs.

- **Methodological Weaknesses:** There were inherent methodological weaknesses in attempting to apply quasi-experimental research designs. Trying to match project and control groups based on important characteristics often proved to be difficult to the point of being impractical.

  Furthermore, extraneous factors were constantly influencing the project and the control groups differently, interfering with the studies and rendering their findings inconclusive despite large expenditures on surveys.

- **Lacked Relevance to Project Management Concerns:** The narrow focus of the design question (i.e., whether impacts were statistically attributable to an intervention), frequently missed many of project management's concerns regarding factors responsible for project success or failure. While useful for answering questions about "accountability" for achieving development results, these methods are typically silent in answering questions about why or how processes have occurred and thus provide few lessons for improving project operations.

Because of methodological and conceptual constraints and because they require massive, expensive data-collection efforts over extended periods of time, such evaluation strategies probably cannot be employed frequently in A.I.D., although they may well be justified in select cases.

While no clear methodology yet exists, there is an emerging consensus that quantitative data on gross trends in impact indicators, if supplemented by appropriate qualitative studies, can satisfactorily address the attribution question by plausibly examining the assumptions and processes linking interventions to successively higher levels of outcomes and by exploring alternative possible explanations for changes.

A related approach involves the development and measurement of more "intermediate" or "leading" indicators that are proxies for the more ultimate, long-range impacts, and that are more easily attributable to the program efforts. For example, in the case of a family planning project, rather than measuring changes in actual fertility rates in the population, an intermediate indicator might monitor the utilization of the project's services in terms of couple years of contraceptive protection distributed or measure contraceptive prevalence rates of the target population. Or in the case of an agricultural project, proxies for production and income impacts might include changes in crop yields, in the volume of commodities passing through markets, or in price movements as indicators of supply changes. These proxy indicators have advantages of being easier to collect, showing "results" well before longer term impacts might appear and closer (and thus more attributable) to the program. However, unless their relationship to the more ultimate impact indicators and goals are clearly established and accepted in the research world, caution needs to be taken in equating the two. For example, in the case of the family planning project, it could be possible that promotion of contraceptives by one program, say a social marketing project, could just be substituting for other public or private sources of contraceptives, and not be affecting overall prevalence or fertility, although intermediate indicators of utilization of the project's contraceptives might be impressive.

(d) Issues in Efficiency Evaluation

Efficiency evaluations are really extensions of impact evaluation in that it compares an intervention's impact with its cost. Efficiency evaluations are useful for comparisons among interventions and thus for resource-allocation decisions. They attempt to answer the question Was investment in this intervention an economical use of scarce resources?

There are two common types of efficiency evaluations in use: cost-benefit analysis (or economic rates of return) and cost-effectiveness analysis. Cost-benefit analysis compares monetized benefits (i.e., net changes or impacts given a monetized value) with costs, whereas cost-effectiveness analysis compares non-monetized benefits with costs, by calculating costs per unit of change. Whereas cost-benefit analysis has the advantage of being able to compare projects with different objectives, it is not always possible to assign a monetary value to desired changes resulting from social projects. For example, how can one value the price of saving a human life? While cost-benefit analyses are appropriate particularly for economic or income-generation interventions, cost-effectiveness assessments can be useful for comparing alternative project approaches aiming at the same program objective. For example, cost-effectiveness indicators for alternative child survival project approaches might compare their average costs per child death averted.

The difficulties in implementing efficiency evaluations are similar to impact assessments; in fact they are really an extension of the same methodology, adding the costs to the analysis of impact. However, there are simpler "proxy" measures being used to compare projects, for example examining costs per project beneficiary or costs in relation to various "intermediate" project performance indicators. For example, alternative family planning project approaches might be compared by looking at their average cost per couple years of protection provided.

(e) Issues in Sustainability Evaluation

The performance and impacts of interventions in the long term after donor involvement ends is of growing concern. However, no clear evaluation methodology has been developed yet to evaluate sustainability. In fact, there is no clear agreement concerning what sustainability means. One view, that sustainability refers to the continuation of project benefits years after donor funding ends, would imply using an impact evaluation methodology, only extended in time beyond the project completion date. More practically or simply, sustainability evaluation methods could involve continued observations of program activities, outputs, and perhaps performance indicators.

Definitions of program sustainability often have a financial aspect, such as the ability of the program to substitute other sources of income for the terminating donor funding. Methods of analysis might include indicators that compare an implementing agency's income generation potential to the expenditures needed to continue the programs' operations and services. Other usages of sustainability, including the institutional capacity of the implementing agency, and the broader issue of environmental sustainability of project activities, will require considerable thought to develop appropriate evaluation methodologies and indicators.

APPENDIX B

Some of the details of the CDIE review of the Agency's decentralized evaluation reports found the following:

- Every year, only about one-fifth of the Agency's project portfolio is evaluated; that translates into approximately one evaluation undertaken sometime during the life of an average five-year project.

- Over two-thirds of the evaluations were interim reports carried out during implementation with the remainder final or end-of-project evaluations. Ex post evaluations were very rare (1%).

- In terms of the primary focus of the evaluations, about two-thirds addressed questions about the project's immediate outputs, with most of the rest focused on project purpose. Only 2% focused on questions about higher order goal achievement.

- There was no systematic or widespread focus in the evaluation reports on issues such as project cost-effectiveness or sustainability, nor did they consistently address special "cross-cutting" interests such as project impacts on the environment or on women.

- While complete data were available on project outputs in nearly one-half of the reports, good data on project purpose achievement were only available in one-fifth of the reports, and were only very rarely available at the goal level (4%).

- Data-collection techniques relied heavily on key informant interviews and, to a lesser extent, on site visits. Little or no use was made of more systematic rapid appraisal techniques such as focus groups or community interviews, informal or formal surveys, or direct observation techniques.

- Over one-half of the evaluation reports called for some form of improvement in the project's monitoring and evaluation system.

- Management sponsoring the evaluations considered less than one-half of the evaluation recommendations to be "actionable" or acceptable for follow-up actions.

- Over one-half of evaluations were conducted by contractors. A.I.D. staff participated as evaluation team members in less than one-third of the evaluations. Similarly, host country staff participated in less than one-third.